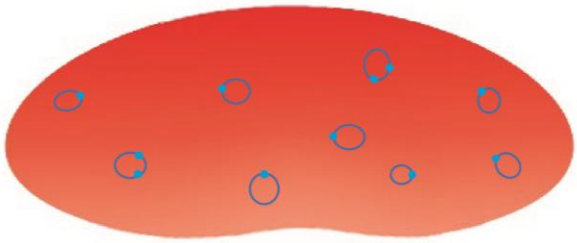
Chapter 1

Introduction and Literature review



1.1 Malaria: historical perspective and current status

Malaria is a life-threatening disease caused by the single-celled protozoan parasites of the genus *Plasmodium*. References to the disease have been made in history dating back to about 2700 BC in a Chinese document, clay tablets from Mesopotamia in 2000 BC, Egyptian papyri in 1570 BC and the Hindu texts dated as far back as the sixth century BC where the symptoms mentioned almost certainly referred to malaria. Malaria fevers were thought to be caused by miasmas rising from marshes for over 2500 years, and it is assumed that the name malaria comes from the Italian term mal'aria, which means spoiled air. Parasite as a causative agent of the disease was learned after Charles Louis Alphonse Laveran, a French army surgeon first noticed and recorded the parasite in the microscopic examination of the blood samples from malaria patients, which he called *Oscillaria malariae* (Cox, 2010). Ronald Ross a British medical officer in Hyderabad, India discovered that *Plasmodium* parasites were transmitted from humans to mosquitoes in 1897 (Cox, 2010). Subsequently, in 1898 Giovanni Battista Grassi and co-workers such as Amico Bignami and Giuseppe Bastianelli demonstrated that the sexual sporogenic cycle of the *Plasmodium* parasite occurs in the *Anopheles* mosquito (Cox, 2010). This led to the discovery of a complete transmission cycle of the parasite.

To date, the genus *Plasmodium* has over 200 species that have been formally described (Sato, 2021). Five species of *Plasmodium* are known to naturally infect humans and cause malaria namely, *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* and *P. knowlesi*. *P. falciparum* is the deadliest among these species, accounting for most of the malaria cases every year. Approximately 229 million malaria cases have been reported worldwide in 2019 (WHO, 2020). The virulence of the species is attributable to the presence of a staggering number of sequence and gene arrays that are involved in pathogenesis and immune evasion (Crabb and Cowman, 2002).

1.1.1 Lifecycle of *Plasmodium*

The complex, multistage, two host lifecycle of malaria occurs between the vector mosquitoes and the vertebrate hosts (Figure 1.1). The sexual stage occurs in an invertebrate definitive host, the female *Anopheles* mosquito while the asexual cycle occurs in a vertebrate host, which is human.

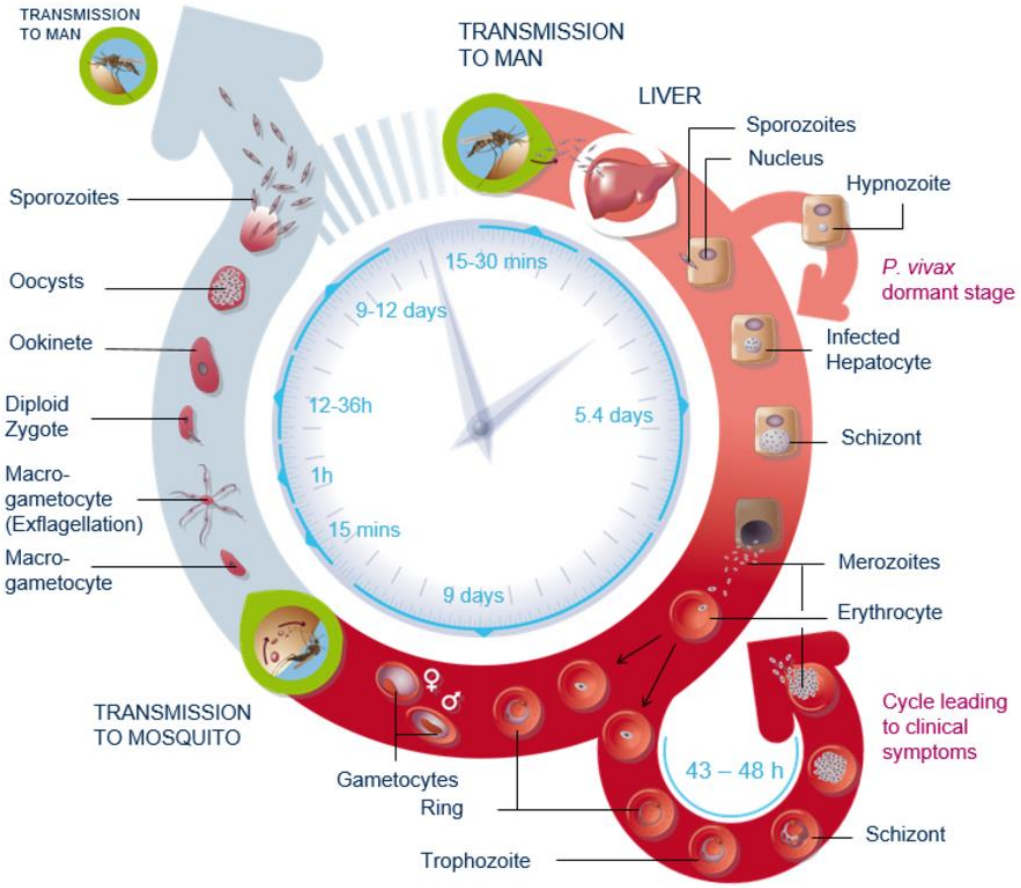


Figure 1.1: Life cycle of *Plasmodium*.

A *Plasmodium* infected mosquito injects sporozoites into the host while taking a blood meal. The sporozoites make their way towards the liver from the site of release, infect hepatocytes and initiate the liver stage infection. The sporozoites replicate inside the hepatocytes to produce thousands of merozoites, which get released into the bloodstream during hepatocyte egress. The merozoites then invade an erythrocyte and initiate schizogony. A parasite inside the erythrocyte develops into a ring-stage which is followed by a mature trophozoite stage after which nuclear division occurs producing a schizont with 16-32 merozoites. The merozoites are released after erythrocyte egress to start a new round of erythrocyte schizogony. A small proportion of merozoites released develop either into a male or a female gametocyte (sexual stages). The male and female gametocytes are ingested by a mosquito during a blood meal and inside the mosquito midgut the male gametocyte exflagellates to form a male gamete which fertilizes the female gamete to produce a zygote that develops into an ookinete. The ookinete penetrates the midgut epithelium and differentiates into an oocyst under the basal lamina. The oocyst grows, ruptures and releases thousands of sporozoites, which then migrate to invade the salivary glands of the mosquito where they are then ready to be injected into a new vertebrate host. **Source**: Medicines for Malaria Venture, https://www.mmv.org/malaria-medicines/parasite-lifecycle.

1.1.2 Classification of malaria

The waste and toxins released after the destruction of red blood cells (RBCs) during the asexual erythrocytic stage are the causative agents of all the clinical symptoms associated with malaria. Based on the severity of the symptoms, malaria can be classified into two categories: uncomplicated and complicated.

1.1.2.1 Uncomplicated malaria

Classical uncomplicated malaria lasts for 6-10 h and it consists of a cold stage, a hot stage and a sweating stage. The cold stage consists of shivering, in the hot stage patient suffers from fever, headache and vomiting, and seizures may develop in young children. Finally, in a sweating stage, the patient sweats while returning to the normal temperature and has a feeling of tiredness.

P. falciparum, *P. vivax* and *P. ovale* with an erythrocytic cycle of 48 h manifest fever after every two days of afebrile interval and therefore it is called tertian malaria. While *P. malariae* with an erythrocytic cycle of 72 h has paroxysms after every three days, thus the infection is termed as quartan malaria. In the countries where the cases of malaria are infrequent, these symptoms may be misdiagnosed as influenza or other common infection. This form of malaria is easily diagnosed in endemic areas and is treatable.

1.1.2.2 Complicated or severe malaria

Serious organ failures or abnormalities in the patient's blood or metabolism exacerbate malaria. The symptoms of severe malaria include cerebral malaria, severe anemia, hemoglobinuria, acute respiratory distress syndrome caused by an inflammatory reaction in the lungs that inhibits oxygen exchange, abnormalities in blood coagulation, cardiovascular collapse leading to low blood pressure, acute kidney failure and hyperparasitemia, where more than 5% of RBCs are infected by the parasite, metabolic acidosis and hypoglycemia. All these symptoms if left untreated usually result in death.

1.1.3 Factors affecting malaria transmission

1.1.3.1 Genetic factors

J B S Haldane speculated that, based upon the genetic makeup, when confronted by a parasitic organism people would have a different risk of dying, to an extent that even if a

gene that protected against the parasite was otherwise harmful, its frequency would increase when a population was exposed to the parasite (Luzzatto, 2012). The hypothesis was apparent in the cases where genetic traits had a protective advantage against malaria. One such trait is the sickle cell trait where heterozygous for the abnormal hemoglobin gene HbS are found to be relatively protected from malaria. The underlying mechanism was proposed to be parasite-triggered sickling of RBCs followed by phagocytosis of the sickled cells (Luzzatto, 2012). This mechanism probably also applies to glucose-6-phosphate dehydrogenase (G6PD) deficient RBCs (Luzzatto, 2012). *P. falciparum* being the primary cause of death in Africa, the sickle cell trait is more commonly found in Africa and persons of African ancestry as compared to other population groups. People with the hemoglobin C trait are another example of such phenomenon, where hemoglobin C is thought to interfere with the ability of the parasite to remodel the host cell cytoskeleton. Furthermore, people who are negative for the Duffy blood group are resistant to *P. vivax* by deleting its chance of invasion into the RBC with the absence of Duffy proteins on the cell's surface which are receptors for the parasite. So, the prevalence of such blood dyscrasias in malaria-endemic areas is believed to provide protection from the disease.

1.1.3.2 Acquired immunity

In areas with a high transmission rate of malaria typically most of Africa of the Sahara, newborns have protective immunity presumably by maternal antibodies. As these antibodies decrease with time, these children become prone to malaria infection and death. But they will have attained a protective semi-immunity, if they survive repeated infections to an older age (2-5 years). Thus, in the high transmission areas young children are a major risk group. In contrast, in areas with a low malaria transmission rate, no protective immunity is attained so, people from all age groups are affected by the disease, which results in epidemics.

1.1.3.3 Pregnancy and malaria

Women with developed immunity against malaria, tend to lose the protection especially during the first or second pregnancies, due to decreased immunity against infectious diseases during pregnancy. The child is at a higher risk of being delivered preterm or with low birth weight, consequently which decreases the chances of survival during the first months of life.

1.1.3.4 Behavioural factors

Poverty in rural malaria-endemic regions brings the inability to afford protective measures such as housing and bed nets that would protect against mosquitoes, accompanied by the inability to recognize malaria and to treat it promptly and properly. Travelers or migrators from non-endemic areas may not follow the preventive measures due to cost, inconvenience or lack of knowledge. Human activities such as borrow pits, standing water in irrigation ditches create breeding sites for mosquito larvae. Other human behavioral factors include work nature which increases the exposure to mosquitoes, raising domestic cattle near households which decreases the transmission rate by providing an alternative source of blood meal.

1.1.3.5 Abiotic factors

There is an association between climatic factors and malaria incidence rate. Both the vector *Anopheles* and the parasite *Plasmodium* are sensitive to temperature changes. Temperatures below 16°-19°C inhibit *P. falciparum* transmission, although *P. vivax* development can occur at temperatures as low as 14.5°-15°C. Temperature above 33°C-39°C may limit the development of both *P. falciparum* and *P. vivax*. A high biting rate of *Anopheles* and a high development rate of *Plasmodium* are expected at 25°C (Kotepui and Kotepui, 2018). Breeding habitats can get flushed away temporarily due to high rainfall, but mosquitoes start breeding as soon as the rain stops. Sometimes, less rainfall interrupts the flow of streams and rivers which causes pooling thus creating a favorable environment for mosquito breeding. Rainfall also affects the transmission of malaria through its indirect effect on humidity. In general, mosquitoes survive better and also become more active under conditions of high humidity. Relative humidity below 60% shortens the lifespan of mosquitoes, which results in low transmission rates (Kotepui and Kotepui, 2018).

1.1.3.6 Type of vectors

Biotic factors include vectors. Mosquitoes of the genus *Anopheles* are the carriers of the malaria parasite. Different species of *Anopheles* mosquitoes have different capacities of transmitting malaria. Mosquitoes from *Anopheles gambiae* group belong to one of the eight morphologically indistinguishable species viz. *A. amharicus, A. arabiensis, A. bwambae, A. coluzzii, A. gambiae sensu stricto, A. melas, A. merus* and *A. quadriannulatus*, which are the most efficient malaria vectors in the world. The anthropophilic behavior of *Anopheles*

is one of the reasons to make them efficient malaria carriers as compared to the other mosquitoes that bite humans and animals equally. While mosquitoes that are strictly zoophilic in behavior cannot be malaria vectors.

Repeated setbacks in the malaria control programs are due to a lack of financial resources in developing malaria-endemic countries which leads to poor public sector healthcare facilities. Whereas the absence of regulatory measures or their proper enforcement encourages private consultation by unlicensed, costly health practitioners and the anarchic prescription and sale of drugs which include counterfeit products. So, this situation poses a major challenge and must be addressed rigorously to make the eradication of malaria successful.

1.1.4 Malaria Vaccine development

The efforts of the World Health Organization (WHO) and partners are underway to develop malaria vaccines with > 75% efficacy against clinical malaria, which will be acceptable for use in all malaria-endemic areas and be licensed by 2030 (Beeson et al., 2019). *P. falciparum* being responsible for most of the malaria cases and deaths has been the major focus of vaccine development.

1.1.4.1 Types of malaria vaccines

Vaccines are broadly grouped according to the malaria parasite life stage they target.

(a) Pre-erythrocytic vaccines

These vaccines target sporozoites and are desirable because they can prevent initial infection and thus prevent clinical illness and malaria transmission. As of 2021 the RTS,S vaccine is the only approved vaccine, known by the brand name Mosquirix, but it has low efficacy. It is based on a virus-like particle that has central repeat and C-terminal epitopes of the major sporozoite surface antigen, circumsporozoite protein (CSP). Recently in 2021, R21/Matrix-M, a candidate malaria vaccine demonstrated efficacy > 77% over 12 months of follow-up in a Phase IIb trial. This vaccine is the first to achieve the WHO's aim of a vaccine with at least 75% efficacy. The vaccine uses a CSP antigen at a higher proportion than the RTS,S vaccine.

(b) Blood-stage vaccines

These types of vaccines generally target the merozoites. They aim to prevent the parasite replication and the development of clinical illness. AMA1, a key invasion protein of merozoites, is the most inspected blood-stage vaccine.

*(*c) Transmission-blocking vaccines

These vaccines target gametocytes and parasite stages in the mosquito midgut. Clinical trials or human population studies have not yet demonstrated the success of this strategy. Some of the leading transmission-blocking vaccine candidates include Pfs230 and Pfs48/45, which are expressed by gametocytes in the human host, and Pfs25, which is exclusively expressed in the mosquito vector (zygote and ookinete stages).

On the basis of the results of vaccine trials, it appears that developing an effective malaria vaccine is exceptionally challenging. One of the major challenges is posed by its vast genetic diversity, especially in the surface antigens that are under continuous selective pressure by the human immune response, and which have also been the main targets for subunit vaccines. The parasite presents different antigens as it changes through several stages of its life cycle. It has also developed a series of strategies to evade the human immune system. To add more to the complexity, it evolves continuously in response to drugs and other malaria interventions through mutation and sexual recombination (Takala and Plowe, 2009). Thus the parasite provides a moving target for the interventions intended against it. The existing vaccines have demonstrated modest efficacy in malaria-endemic areas, which indicates that substantial challenges still remain, to achieve a highly efficacious and durable vaccine but there is great hope for the future.

1.1.5 Antimalarial drugs

The two important currently used antimalarial drugs, artemisinin and quinine have their roots in the plants whose medicinal values were noted many centuries ago. Artemisinin was derived from *Artemisia annua* L. in China in the fourth century (Wright et al., 2010) and quinine was extracted from the plants of the genus *Cinchona* (quina in china) in South America in the seventeenth century (Achan et al., 2011). The bark of this tree stores several biologically active alkaloidal natural products, of which the four most abundant ones are the quinolone family compounds quinine, quinidine, cinchonidine, and cinchonine. Until the

mid1800's the powder prepared from the bark of this tree, Jesuit's powder was used for the treatment. In 1820, Pierre Joseph Pelletier and Joseph Caventou isolated the active ingredient quinine (Achan et al., 2011) and made malaria the first disease for which pure compound was available for treatment. Efforts to synthesize quinine were started in 1856, but until 1944 this goal was not accomplished and it has never been achieved on a commercially economic scale. Quinine shortages during World War I were accompanied by the need to find new antimalarial drugs. Subsequently, during World War II chloroquine (CQ) was found to rapidly control the clinical symptoms of malaria with minimal toxicity and was approved by the FDA in October 1949. CQ served as the mainstay for malaria therapy for at least four decades (Deshpande and Kuppast, 2016) until the emergence of CQ resistant *P. falciparum* and *P. vivax* strains which rendered it less useful.

1.1.5.1 Classification of antimalarials

Antimalarial drugs can be categorized into various classes according to their mode of action.

(a) Quinolones and the related compounds

These compounds act by blocking the detoxification of heme in the parasite. The class includes quinine, quinidine, CQ, amodiaquine, mefloquine, piperaquine and primaquine, and an aza-acridine which is structurally related to the quinoline antimalarials (Rawe and McDonnell, 2020).

(b) Antifolates

They act by disrupting the metabolic pathways that demand one-carbon moieties supplied by the B9 folate vitamins, which they mimic (Visentin et al., 2012). Pyrimethamine, proguanil and the sulfa drugs such as sulfonamide, sulfadoxine, sulfone and dapsone are the principal antifolate drugs used against malaria.

(c) Artemisinin and its derivatives

The mechanism of action of these compounds is understood to involve the heme-mediated decomposition of the endoperoxide bridge in the drug to produce carbon-centered free radicals, which generate oxidative stress in the parasite. The class comprises the parent compound artemisinin, *dihydroartemisinin*, *artemether*, *artemotil* (*arteether*) and *artesunate*.

CQ, sulfadoxine-pyrimethamine and artemisinin are the frontline drugs to treat malaria. Due to the persistent threat of malaria evolving resistance to the available drugs, artemisinin derivatives are used in combination with a 'partner drug' from a different drug class and the resulting drug combinations are known as artemisinin-based combination therapies (ACTs).

Some of the ACTs are artemether-lumefantrine, artesunate-amodiaquine, artesunate-sulfadoxine-pyrimethamine, artesunate-mefloquine and dihydroartemisinin-piperaquine (Price and Douglas, 2009). At present, ACTs remain the most effective method to treat the resistant parasite but again with the reports of multidrug resistance, besides artemisinin (partial) resistance and partner drug resistance in four Greater Mekong Subregion (GMS) countries and the countries outside GMS (WHO, 2018). Consequently, along with understanding the emergence and selection of drug-resistant strains, which is necessary to stop the constant battle, it is imperative to discover new antimalarial drugs, before the massive death history repeats itself.

1.2 Plasmepsins: the *Plasmodium* aspartic proteases

Plasmepsins (PMs), the aspartic proteases from *Plasmodium*, play an important role in each stage of the parasite's development. The first sign of involvement of acidic proteases in hemoglobin digestion was found nearly 75 years ago during the characterization of cell-free extracts from *P. gallinaceum* (Moulder and Evans Jr, 1946). Thereafter, hemoglobin degrading acidic proteases were reported in numerous species of *Plasmodium* (Nasamu et al., 2020), some of which were demonstrated to be blocked by canonical aspartic protease inhibitor, pepstatin A (Aissi et al., 1983; Sato et al., 1987). Participation of several enzymes in hemoglobin digestion was suspected and was confirmed by the presence of aspartic hemoglobinase I and II in *P. falciparum* (Francis et al., 1994). These two initially identified proteases were systematically renamed such that, plasmepsin I replaced aspartic hemoglobinase I and *Plasmodium* aspartic proteinase G and plasmepsin II replaced aspartic hemoglobinase II and *Plasmodium* aspartic proteinase D (Hill et al., 1994). The name plasmepsin (PM) comes from the organism *Plasmodium* and a common aspartic protease pepsin which has a similar molecular structure.

The discussion is focused on PMs from *P. falciparum*. The completion of the genome sequencing effort of *P. falciparum* revealed the presence of ten aspartic protease genes, PM I, II, IV-X and histo aspartic protease (HAP) (Coombs et al., 2001). The genes for PM I, II,

HAP, IV, VIII and IX are located on chromosome 14 while for PM V, VI, VII and X they are situated on chromosomes 13, 3, 10 and 8 respectively (Nair and Singh, 2016). Localization studies placed PM I, II, HAP and IV from *P. falciparum* in the parasite's food vacuole (Banerjee et al., 2002). Spread across 16 kb of chromosome 14 they share 50-70% of amino acid identity. In the species other than *P. falciparum* and in other related primate infecting species, these proteases are represented by a single protease, called PM IV in *Plasmodium* and aspartic protease 1 in the related apicomplexan *Toxoplasma gondii* (Shea et al., 2007). The all four enzymes from *P. falciparum* were found to be important for hemoglobin degradation during the intraerythrocytic stage of the parasite (Nasamu et al., 2020). The enzymes PM I and II are 73% identical to each other and are about 35% homologous to mammalian renin and cathepsin D (Francis et al., 1997). PM I and PM II initiate the process of hemoglobin degradation by cleaving the native molecule in a highly conserved hinge region (Banerjee et al., 2002), which is involved in maintaining the integrity of the molecule as it binds oxygen (Francis et al., 1994). They perform the first cleavage of the molecule between Phe33 and Leu34 of the α-chain, resulting in protein unfolding and release of the heme moiety (Gupta et al., 2010).

1.2.1 Biosynthesis and maturation

The genes for PM I and PM II encode around 51 kDa proteins (Francis et al., 1997) (Figure 1.2). Typically aspartic proteases are synthesized as inactive zymogens as pre-proproteins in order to prevent unwarranted degradation of proteins (Koelsch et al., 1994). The prosegments, characteristically comprised of 50 amino acids are removed during protein maturation. But in the case of PM I and PM II these are 123 and 124 amino acids long (~70 residues longer) respectively. Cleavage of these prosegments yields mature forms of approximately 37 kDa. In both PM I and II, 65 amino acids upstream of the cleavage site is a hydrophobic stretch of 21 amino acids which has been predicted to serve as a signal anchor sequence. Several studies including those involving hydropathy plots of proPMs and on their biosynthesis have categorized them as type II integral membrane proteins that traverse the secretory pathway before cleavage to the soluble form. There is also an agreement from the positive inside rule. The NH_2-terminal of the proPM I and proPM II (overall prosegments of these enzymes) bear positively charged amino acid residues, thus makes the membrane bound section of the proteins. Whilst, the mature forms of the enzymes are derived from the carboxyl terminus of the proproteins (Francis et al., 1997).

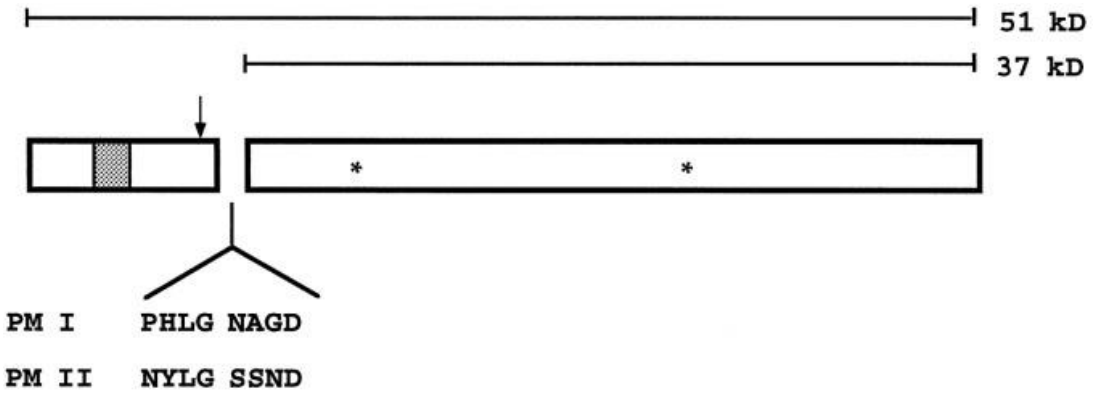


Figure 1.2: Schematic representation of PM I and PM II.

The preforms of both the enzymes are approximately 51 kDa proteins. There is a 21 amino acid hydrophobic stretch after 37 amino acids from the initiator methionine in proPM I and 38 amino acids from the initiator methionine in proPM II, which is predicted to be a signal anchor sequence. Both the proenzymes are processed to mature PMs of approximately 37 kDa. The cleavage site and the 4 adjacent amino acids (P4, P3, P2, P1, P1', P2', P3', P4') of both the proteins are depicted. The active site aspartic acids (Asp34 and Asp214) of both the PMs are denoted by asterisks (*). Source: modified from (Francis et al., 1997).

The vacuolar PMs, PM I-IV are synthesized following a circuitous route (Figure 1.3). The enzymes synthesized as transmembrane zymogens are inserted as type II membrane proteins into the endoplasmic reticulum (ER) or the nuclear envelope. Transport vesicles containing the enzymes bud from ER exit sites, traffic through the secretory pathway and reach the parasite surface. These vesicles migrate to the cytostome, an invagination by means of which parasite ingests hemoglobin and that extends from the parasite plasma membrane to the parasitophorous vacuolar membrane (a red cell-derived structure with which the parasite surrounds itself as it invades). The vesicles merge with the outer membrane of the cytostomal vacuole and this places the proenzymes in the space between the two vacuolar membranes. Internalizing the substrate hemoglobin, the double-membrane vesicle pinches off from the cytostome, travels to the food vacuole and its outer membrane merges with that of the food vacuole. This event leaves the proPM anchored to the membrane of the food vacuole. Ensuing this, the cleavage of the proregions releases the mature PM enzymes in the vacuolar lumen (Klemba et al., 2004) which are capable of hemoglobin degradation.

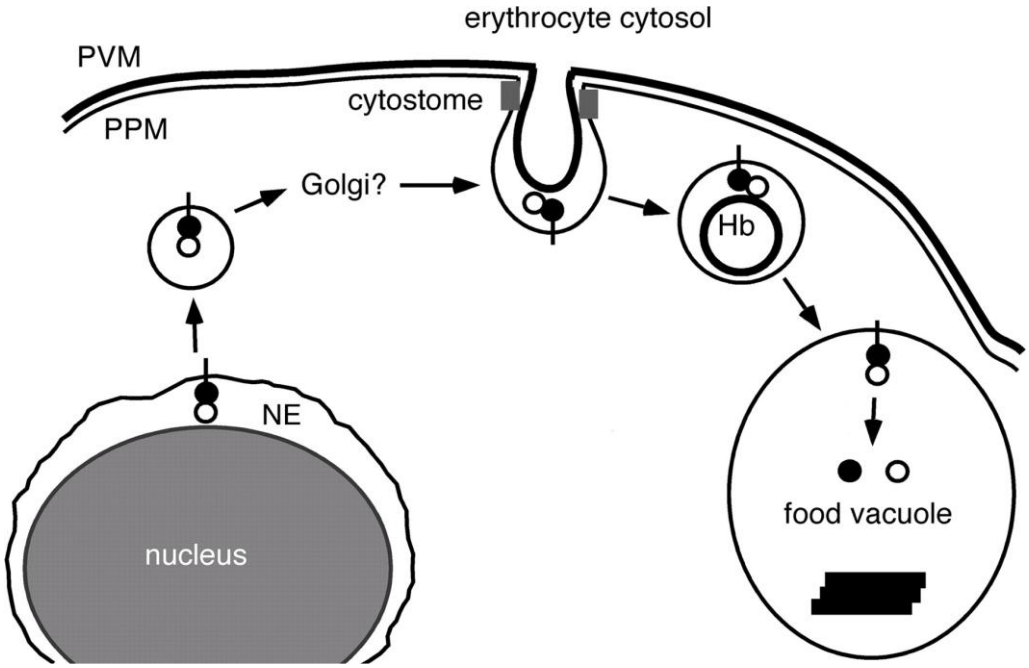


Figure 1.3: Model showing route of vacuolar proPMs trafficking.

Trafficking of the food vacuole PMs is illustrated using the route proposed for GFP tagged proPM II (proPM II-GFP). proPM II-GFP is incorporated as a type II membrane protein into the ER membrane. The transport vesicles containing proPM-GFP bud from the ER and travel to the cytostome (possibly via a golgi-like compatment). The vesicles merge with the outer membrane of the cytostome, which is topologically contiguous with the parasite plasma membrane. This event places proPM II-GFP in the space between the two vacuole membranes. The double-membraned vesicle containing the ingested hemoglobin pinches off from the cytostome, and travels and merges with the outer membrane of the food vacuole. This leaves proPM II-GFP anchored in the food vacuole membrane. The proregion of the enzyme and the GFP tag are removed to yield mPM II. Black circles: PM II; White circles: GFP; PVM: parasitophorous vacuole membrane; PPM: parasite plasma membrane; Hb: hemoglobin. Source: (Klemba et al., 2004).

1.2.2 Activation mechanism of vacuolar PMs

The mature PMs are formed by the removal of the prosegments during activation. Various insights have been made from extensive biochemical and structural studies of vacuolar PMs, but the mechanism of activation is still unclear. The prosegment of a vacuolar PM is comprised of an initial β-strand followed by two α-helices (H1 and H2) and a pro-mature region (a segment joining the helix H2 to the mature part of the protein) that harbors a Tyr-Asp loop and a cleavage site. The two helices H1 and H2 interact exclusively with the C-terminal domain of the enzyme. Such interactions render the active site cleft more open in the zymogen than in the enzyme, thus separating the catalytic aspartates (Asp and His in HAP) in the active site of the enzyme (Khan et al., 1999). This method of inactivation in

vacuolar PMs is different from that observed in zymogens of gastric aspartic proteases. In pepsinogen e.g., a lysine residue in the prosegment neutralizes the charge repulsion between the two catalytic aspartates at neutral pH, which keeps the enzyme in an inactive zymogen form. At low pH, the salt bridge interactions that stabilize the prosegment position across the active site of pepsin are assumed to be disrupted. Following the release of the prosegment, this event opens the substrate-binding cleft. This leads to the formation of mature pepsin from pepsinogen after auto-activation that is catalyzed by the active site of the same enzyme molecule.

The possibility of vacuolar PMs exhibiting an activation mechanism similar to that of pepsinogen was explored (Rathore et al., 2021). The study showed that the vacuolar proPMs are present in an S-shaped dimer form (Figure 1.4). The interactions between the Tyr-Asp loops in the prosegments of the monomers are essential for the stability of the dimer. Also, the formation of this dimer is facilitated by widespread stacking interactions between hydrophobic residues of helix H2 and a loop L1 (loop present in the mature part of the enzyme) in each of the two monomers. Under acidic conditions, the disruption of the Tyr-Asp loop interactions breaks the inter-residue hydrogen bond and moves the loop L1 and helix H2 downwards. This further extends the pro-mature region which causes movement of the monomers that leads to weakening of their interactions at the dimer interface, facilitating dissociation of the dimer into monomers. The C-terminal part and the pro-mature region of the prosegment are understood to be more flexible parts of the structure as compared to the others. Therefore further under acidic conditions, this segment of the polypeptide in the proPM monomer is thought to unfold and become susceptible to hydrolytic cleavage. This unfolding of the prosegment of the enzyme under acidic conditions was found to be reversible. The unfolding of the prosegments places the two catalytic aspartates in close proximity, forming a competent active site. Such an active site was found to be accessible to the substrate, which suggested that the enzyme is capable of cleaving its substrate peptide. As opposed to pepsinogen, as the cleavage site in the vacuolar proPMs is located distantly from the active site, it was suggested to be unlikely that the proPMs perform self-cleavage of their own prosegment. It was therefore suggested that the proPMs act transiently and cleave prosegment of other unfolded proPM molecules. The first molecule of mature enzyme formed after an initial cleavage event is thought to rapidly convert all the proenzymes into their mature forms and the cycle is suggested to last until all protein is activated.

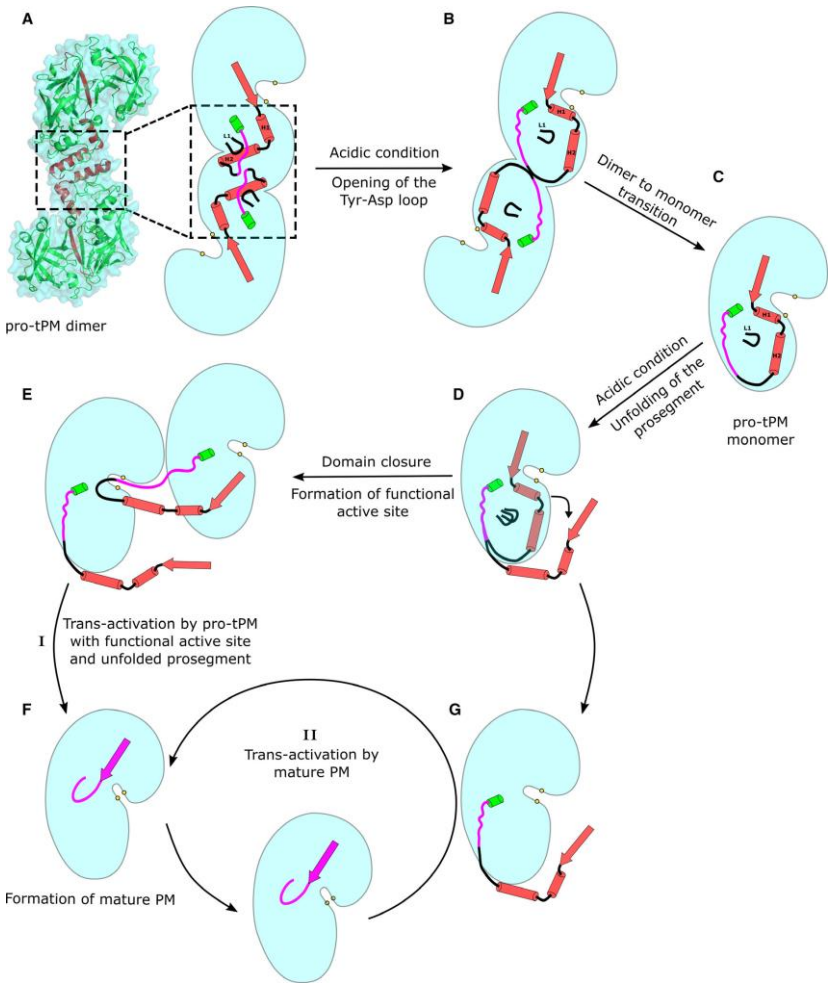


Figure 1.4: Schematic representation of mechanism of trans-activation of vacuolar proPMs.

Mechanism of activation of the vacuolar proPMs is illustrated using the mechanism of activation proposed for the proPMs with truncated prosegment (pro-tPMs). The prosegment in the PMs is comprised of an initial β-strand followed by two α-helices (H1 and H2) and a pro-mature region that harbors a Tyr-Asp loop. The pro-tPMs exist as an S-shaped dimer that is stabilized by interactions between the Tyr-Asp loop of the monomers and by inter-residue interactions between the helix H2 and a loop L1 (loop present in the mature part of the enzyme) of the two monomers (A). Under acidic conditions, the weakening of the Tyr-Asp loop interactions moves the loop L1 and the helix H2 downwards (B), which extends the pro-mature region and disrupts the interactions between the monomers at the dimer interface, facilitating dissociation of the dimer into monomers (C). Further under acidic conditions, the prosegment unfolds (D) which brings the two catalytic aspartates in close proximity, forming a competent active site in the monomers. Following this a pro-tPM acts transiently to cleave the prosegment of another unfolded proPM molecule to produce a mature PM (E). The first molecule of the mature enzyme formed (E) after this initial cleavage event rapidly converts all the proenzymes (G) into their mature forms and the process lasts until all the protein is activated. The helices H1 and H2 are shown as cartoons in red and the pro-mature region is represented in magenta color. The region in cyan denotes the surface of the different forms of PMs. Source: (Rathore et al., 2021).

In vitro studies have reported auto-activation of truncated zymogen of PM II at low pH, with cleavage occurring at a different site (+2 and +12 upstream) as compared to that of the native cleavage site (Tyas et al., 1999). The cleavage site has been observed to vary depending on the reaction conditions. It has been proposed that the prosegment unfolds and allows the successive cleavage across its length during the activation process (Rathore et al., 2021). An involvement of maturase in the activation of the vacuolar PMs has also been suggested. A low pH has either been suggested to be necessary to assume the extended conformation suitable for proteolytic cleavage or it has been suggested to be required if the supposed enzyme, maturase itself has an acidic pH optimum (Khan et al., 1999). But no further studies have confirmed any such enzyme-assisted activation of PMs.

1.2.3 Structure and mechanism of action

PM I, II, IV are known to be classical aspartic proteases, with pepsin-like enzymes catalytic site formed at the junction of two domains containing an aspartic acid residue each (Xiao et al., 2007). Whereas one of the active site aspartic acid residues is replaced by histidine in the case of HAP (Berry et al., 1999). The active sites of PM I, PM II, and PM IV contain two catalytic aspartic acid residues, Asp34 and Asp214, whereas Asp34 is replaced by a histidine residue in HAP.

Several studies of PMs bound with peptidic inhibitors have been performed to elucidate the catalytic mechanism of action and delineate the active site situated in the large cleft between the N- and C- terminal domains of the enzymes (Bhaumik et al., 2012). In general, the catalytic aspartic acid residues participate in acid-base catalysis and activate the water molecule through a tetrahedral intermediate (Dunn, 2002) (Figure 1.5). This starts when the first residue Asp34 is protonated and the second residue Asp214 remains negatively charged. The negatively charged Asp214 acts as a general base to remove one proton from the water molecule trapped between these two residues. The activated nucleophile attacks the carbonyl carbon of a scissile peptide bond to create a tetrahedral oxyanion intermediate. Meanwhile, Asp34 donates a proton to the carbonyl oxygen atom of the scissile bond, which destabilizes the intermediate. Then hydrogen from Asp214 is transferred to the nitrogen of the scissile bond by inversion of configuration around the nitrogen atom. Following this, the C-N bonds break forming two products. Asp214 in its negatively charged form positions itself for the next round of catalysis.

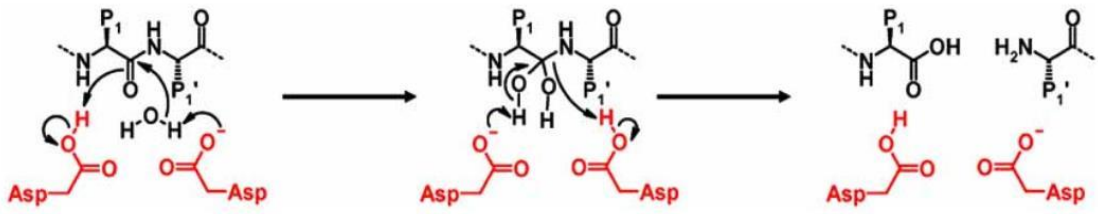


Figure 1.5: Catalytic mechanism of aspartic proteases.

One of the catalytic aspartic acid residues in an aspartic protease activates a water molecule by abstracting a proton, enabling it to perform a nucleophilic attack on the carbonyl carbon of the scissile bond of the substrate. This results in the formation of a tetrahedral oxyanion intermediate which is stabilized by hydrogen-bonding with the second aspartic acid residue. The rearrangement of this intermediate leads to protonation of the scissile amide, eventually splitting the substrate peptide into two product peptides. Catalytic aspartic acid residues (Asp34 and Asp214) are shown in red. P1: substrate residue N-terminal of the scissile bond; P1': substrate residue C-terminal of the scissile bond. Source: Creative diagnostics https://www.creative-diagnostics.com/aspartic-proteases-and-regulators.htm.

1.2.4 PMs as antiplasmodial drug targets

Hemoglobin degradation is one of the most vital metabolic processes for the survival of the *Plasmodium* parasite and the process is initiated by the two vacuolar PMs, PM I and PM II in *P. falciparum*. They share only about 35% sequence homology with human aspartic protease, cathepsin D (Silva et al., 1996). Inhibitors designed against them have demonstrated inhibition of *P. falciparum* growth *in vitro* (Silva et al., 1996). In addition, there is also the presence of sound X-ray crystallographic data on PMs. Altogether, this strongly suggests PM I and PM II as the potential targets for developing novel antiplasmodial drugs.

1.3 Drug discovery from plants

Plants have been a source of food for energy and medicine for the maintenance of health since prehistoric times. While the ancient man was struggling to treat afflictions, in the absence of any information on plants concerning their property, usage, etc. the journey from food to becoming medicine was completely “hit and trial”. Gradually the treatment became to be based on experience, and experience collected over the years resulted in treatment founded on explicatory facts. Pieces of evidence for use of plants for treatment have been

found on Sumerian clay tablets from Nagpur, which were written in 5000-3000 BCE. The inscriptions mention 12 drug preparation recipes which refer to over 250 various plants, including some plants containing psychoactive alkaloids such as poppy, henbane and mandrake (Petrovska, 2012). Originally shared as an oral tradition, Ayurveda, one of the world's oldest holistic healing systems that provides information on the usage of herbs for the treatment of diseases, originated in India more than 5000 years ago. Siddha, a system of traditional medicine originated in ancient Thamilakam in South India (10,000-4,000 BCE) (Devarajan and Mohan Maruga Raja, 2017). The use of plants as medicine is well documented in the Chinese book "Pen T'Sao" (circa 2500 BCE), written by Emperor Shen Nung which lists 365 drugs (dried parts of medicinal plants).

Thus, different medical traditions aroused from various early civilizations namely Ancient Babylonian Medicine from Mesopotamia, Ancient Egyptian Medicine from Egypt, Ayurveda and Siddha from India and Traditional Chinese Medicine from China, which were then inherited by Ancient Greek Medicine. The roots of modern or western medicine can be traced back to Hippocrates of Kos (460-370BC), a Greek physician referred to as the "Father of Medicine" who marked a stage in medicine where the disease was coming to be regarded as a natural rather than a supernatural phenomenon. Early roman contributions such as Dioscorides's "De Materia Medica" (40-90) also influenced the development of modern medicine. The book described 657 drugs of plant origin including the procedure for medicinal preparation and their therapeutic effects, and was famous as the world bible until the sixteenth century.

In the late eighth century, while separating the apothecary and the physician, Arabs introduced the concept of pharmacies by establishing the first privately owned drug stores. Iba Sina – The "Persian Galen" (Avicenna: western world) (980-1037), wrote "The canon of medicine", a medical encyclopedia, which was widely studied as a medical text in universities during the medieval period and was used until 1650.

In the nineteenth century, the application of chemical analysis radically altered the place of plant medicines. Different classes of pharmacologically active metabolites were discovered and isolated from plants. With the progress in purification techniques, the isolated and purified active compounds rather than the whole plants began to be preferred. During the early 1980s, with advancements in synthetic organic chemistry, structural biology and computer technology, the attention shifted from random searches from natural products to

a rational computational archetype for drug discovery, "rational drug design" or sometimes referred to as "computer-aided drug-design". Rational drug design involves designing the molecules that are complementary in shape and charge to the biomolecular target, to create interaction and produce the intended effect. Although there were a few successes from the "rational drug design" approach even before 20 years of its arrival, in the 1990s there was a return made to empirical methods integrated with sophisticated techniques which accelerated the exploration of nature to find out novel drugs.

In the twentieth century, both the approaches: the "empirical approach" and the "rational drug design" are simultaneously applied to discover or construct valuable prophylactic or therapeutic agents, but it is undeniable fact that most of the synthetic compounds are inspired by nature. Based on the analysis performed by Newman and Cragg (2016), out of 1562 drugs approved by FDA from 1981-2014, 67 (4%) were unaltered natural products, 9 (1%) were botanical drugs (defined mixture), 320 (21%) were natural product derivatives, 172 (11%) were synthetic products which were natural product mimics (more natural product like or direct competitive inhibitors of natural substrates), 61 (4%) were synthetic drugs with natural product pharmacophore and 162 (10%) were synthetic drugs with natural product pharmacophore which were also natural product mimics. So, altogether the products of natural origin account for a total of 791 FDA approved drugs.

1.3.1 Traditional wisdom: Ayurveda, Siddha, Unani

India being the land of origin of many traditional philosophies has given birth to many traditional systems of medicine while some traveled from the other parts of the world and took firm roots in the soil. Ayurveda, the oldest medical complex system about healthy lifestyle principles was created in India. The Sanskrit word Ayurveda is derived from two roots: "Ayur" which means life and "Veda" which means science. So, Ayurveda literally translates to "the science of life". The origin of its teachings dates back to the ancient scriptures (Vedas) which came into existence due to the inner realization of illuminated sages (rishis). The earliest writing "Rig Veda" the collection of Hindu sacred verses led to a system of healthcare known as Ayurvedic medicine which beliefs in holistic management of health and disease. The other main Ayurvedic classics are Atharvaveda (around 1200 BC), Charak Samhita and Sushrut Samhita (1000-500 BC), which together describe over 700 plants with their classification, pharmacological and therapeutic properties. "The legacy of Charaka" (2003) and "The legacy of Susruta" (2007) by M. S. Valiathan are scholarly

descriptions of original texts, with commentary from modern medicine and science standpoint besides some glimpses of ancient wisdom. The Ayurvedic Formulary of India is the first official document on Ayurveda and its latest edition details 985 formulations. The Ayurvedic Pharmacopoeia comprises more than 1,200 species of plants, nearly 100 minerals and over 100 animal products (Sekar and Mariappan, 2008). It has 645 monographs on single drugs, 202 on formulations.

The Siddha system of medicine is believed to have originated during the Indus civilization and then was popularized between 2500 and 1700 BCE (Gopal, 2020). This system has its roots intertwined with the ancient Tamil civilization. The word Siddha comes from the Tamil word "Siddhi" which means achievement. According to a traditional belief, Lord Shiva unfolded the knowledge of the Siddha medicine system to his concert Parvati, who handed it down to Nandhidevar and he in turn to the 18 premier scholars called siddhars. Agastyar or Agastya, one among the 18 siddhars is believed as the father of Siddha medicine. The materia medica of the Siddha system depends largely on drugs of metal and mineral origin, in contrast to that of Ayurveda of an earlier period, which was mainly based upon drugs of vegetable origin (Ravishankar and Shukla, 2008). The Siddha Pharmacopoeia of India consists of 139 monographs on single drugs and the Siddha Formulary of India describes 399 formulations.

The Unani system of medicine, owes its immediate origin to the ancient Greece which was known as Yunan. Since Greeks adopted medicine from Egypt, the roots of this system go to the two sister civilizations, Egyptian and Mesopotamian. The Greek period of the Unani medicine began with the great scholar of medicine, Asclepius (1200 BC) (Department of Ayush, 2016). After crossing through many countries, the Unani medicine was introduced in India around the eighth century by Arabs and Iranians. The drugs employed in this system are mainly derived from plants while some are of animal and mineral origin (Ravishankar and Shukla, 2008). The Unani pharmacopeia of India comprises 298 monographs on single drugs and 150 monographs on formulations, while the National Formulary of Unani Medicine describes 1229 formulations.

Although medical pluralism is widely practiced in India, Ayurveda still remains dominant even as compared to modern medicine particularly for the treatment of chronic disease conditions (Patwardhan et al., 2004).

1.3.2 Ethnopharmacological approach: research from clinics to laboratories

It is estimated that approximately 80% of the world's population living in the developing countries rely on traditional medicinal remedies majority of which are derived from plants (Ekor, 2014). The great medicinal system of Ayurveda is a living tradition in practice in India even today. One of the approaches to utilize the knowledge embedded in the traditional medicinal system to aid the discovery of new drugs, is ethnopharmacology. This approach involves field explorations of indigenous medical knowledge and biodiversity that can accelerate the search for newer, safer, effective as well as affordable medicines. The ethnopharmacology knowledge and experiential base allow the research from 'Clinics to Laboratories', in other words, clinical experiences, observations or available data becomes a starting point of the drug discovery process whereas, in the conventional drug research, it comes at the end. So, the ethnopharmacological approach is a true Reverse Pharmacology Approach (Patwardhan et al., 2004).

Using the ethnopharmacological approach, a large number of currently prescribed drugs for curing a wide variety of ailments including malaria, cardiovascular diseases, cancer, etc. have originated from natural resources. Some of the prominent examples include:

Isolation of vasicine (Respiratory stimulant) from *Adhatoda vasica* Nees (*Acanthaceae*), santonin (Anthelminthic) from *Artemisia maritime* L. (*Asteraceae*), atropine (Anticholinergic) from *Atropa belladonna* L. (*Solanaceae*) and camptothecin (Anticarcinogenic) from *Camptotheca acuminate* Decne. (*Cornaceae*). Worthy of mention are also the two most important antimalarial drugs, quinine from *Cinchona ledgeriana* Moens ex. Trimen (*Rubiaceae*) and artemisinin from *Artemisia annua* L. (*Asteraceae*).

The presence of phytochemicals with diverse activities, the ease of availability and the long history of dependence on plants for treating diseases motivate us to explore the traditional knowledge to discover the much-needed new antimalarial agents from plants.

Rationale, Scope and Objectives of the Thesis

Malaria, the disease caused by the protozoan parasite *Plasmodium* is a leading cause of death in many developing countries. *P. falciparum* is the most virulent species among all the five species of *Plasmodium* that infect humans. Diagnosis and treatment of the disease are the most essential interventions to fight against it. But the major challenges to the treatment are posed by the emergence of the *Plasmodium* strains that are resistant to the frontline antimalarial drugs namely, CQ, sulfadoxine-pyrimethamine and artemisinin, and ACTs. So, the situation calls for the discovery of new antiplasmodial agents before the disease repeats its massive death history.

Being the leading cause of malaria-related death, the search for antiplasmodial agents against *P. falciparum* becomes necessary. The food vacuole enzymes PM I and PM II from *P. falciparum* that initiate the process of hemoglobin degradation are essential for the survival of the parasite. The sequence identity between these enzymes and human aspartic protease cathepsin D is only about 35%. The inhibitors targeted against them have demonstrated inhibition of *P. falciparum* growth *in vitro*. Thus, PM I and PM II are potential targets for new antiplasmodial drugs.

Plants produce myriads of molecules with diverse pharmacological properties, indeed many of the molecules still remain unexplored. Since antiquity, people have relied on plants for medicine. The traditional Indian medicinal system, Ayurveda, and the traditional practices involving the usage of plants as medicine provide information on the various properties of the indigenous plants. The ethnopharmacology approach can be a successful tool for utilizing the knowledge embedded in traditional medicinal systems to discover the inhibitors of PM I and PM II from plants.

In light of the above, the present study aimed at investigating the indigenous plants to find out antiplasmodial agents that are inhibitors of *P. falciparum* PM I and PM II. The selection of plants was done using the ethnopharmacological approach. To meet the requirement of the study the enzymes PM I and PM II were expressed recombinantly. Recombinant expression of PMs and achieving their active forms have been challenging to date. So, efforts were done to establish an expression system to achieve the enzymes in their active forms. The study was extended to explore the interactions between the putative inhibitors

and the target enzymes and gain insight into their mechanism of action, using *in silico* approach.

To achieve the aforementioned aim, the objectives of the study were identified as the following:

1. Screening of medicinal plant extracts for antiplasmodial activity.
2. Cloning and expression of PM I and PM II from *P. falciparum*.
3. Screening of the antiplasmodial plant extracts to find out putative PM I and PM II inhibitors.
4. *In silico* studies on PM I and PM II and their putative inhibitors from the plant extracts.